Identification of At-Risk Students Using Network Analysis

Vrushank Ahire
2022CSB1002

April 25, 2024

Abstract

In the realm of educational institutions, ensuring student success and mitigating potential academic risks are of paramount importance. This paper presents a novel approach to predictive analysis and identification of at-risk students using network analysis techniques. By leveraging the intricate connections and relationships between students, we aim to gain valuable insights into their academic performance and potential challenges. Through the construction of a student network and the application of centrality measures, we explore methods to predict academic outcomes and identify individuals who may require additional support or interventions. The proposed methodology has the potential to revolutionize the way educational institutions approach student support, enabling proactive measures and targeted interventions to foster academic success.

1 Introduction

Student success is a multifaceted concept that extends beyond mere academic performance. It encompasses factors such as social integration, emotional well-being, and overall engagement with the educational environment. Identifying students who may be at risk of academic difficulties or social isolation is crucial for educational institutions to provide timely support and interventions. Traditional methods of assessing student performance, such as grades and test scores, often fail to capture the underlying dynamics and interconnectedness within the student community.

Network analysis, a powerful tool from the field of graph theory, offers a novel approach to understanding the complex relationships and interactions among students. By representing students as nodes and their connections as edges, we can construct a student network that captures the intricate web of social ties and academic collaborations. This network serves as a rich data source, enabling us to analyze various centrality measures, identify potential clusters, and uncover patterns that may contribute to academic success or failure.

2 Importance of Network Analysis in Education

The application of network analysis in the educational domain holds significant potential for several reasons:

1. **Early Identification of At-Risk Students**: By analyzing the network structure and centrality measures, we can identify students who may be socially isolated or have limited connections within the student community. These individuals may be at higher risk of academic difficulties or disengagement, allowing educators to intervene proactively.

2. **Predictive Analysis of Academic Performance**: Leveraging the network data and incorporating additional student attributes, we can develop predictive models to estimate academic performance, such as grades or exam scores. This approach provides a holistic understanding of the factors influencing student success.
3. **Targeted Interventions and Support Strategies**: By identifying patterns and clusters within the student network, educators can design tailored interventions and support strategies. These may include fostering collaborative learning environments, facilitating peer mentorship programs, or providing personalized academic guidance.
4. **Fostering Social Integration and Inclusion**: Network analysis can reveal potential social divisions or disconnected subgroups within the student population. This information can guide efforts to promote social cohesion, foster inclusive learning environments, and facilitate cross-cultural exchange.

3 Methodology

The proposed methodology involves several key steps:

1. **Data Collection**: Gather data on student connections, attributes, and academic performance. This may include survey responses, course enrollments, extracurricular activities, and academic records.
2. **Network Construction**: Construct a directed graph representing the student network, where nodes represent individual students, and edges represent connections or relationships between them. These connections can be derived from various sources, such as shared courses, study groups, or self-reported friendships.
3. **Centrality Measures**: Compute various centrality measures, including degree centrality, clustering coefficient, and betweenness centrality, to quantify the importance and connectivity of each student within the network.
4. **Feature Engineering**: Extract relevant features from the network structure, centrality measures, and student attributes for use in predictive modeling and risk assessment.
5. **Predictive Modeling**: Employ machine learning techniques, such as regression or classification algorithms, to predict academic performance or outcomes based on the extracted features.
6. **At-Risk Student Identification**: Analyze the network structure and centrality measures to identify students who may be socially isolated or have low connectivity within the network. These individuals may be at higher risk of academic difficulties and require targeted interventions.
7. **Validation and Evaluation**: Evaluate the predictive performance of the models using appropriate metrics, such as accuracy, precision, recall, or mean squared error. Assess the effectiveness of the at-risk student identification process through feedback from educators and student support services.

4 Mathematical Analysis

Network analysis relies on various mathematical concepts and measures to quantify the importance and connectivity of nodes within the network. In the context of student networks, we focus on the following centrality measures:

4.1 Degree Centrality

Degree centrality measures the number of connections a student has within the network. For a directed graph, we have in-degree centrality and out-degree centrality. The degree centrality of a node i is defined as:

$$C_D(i) = \frac{k_i}{n-1} \tag{1}$$

where k_i is the degree (number of connections) of node i, and n is the total number of nodes in the network. The degree centrality is normalized by dividing by $n-1$, the maximum possible degree.

4.2 Clustering Coefficient

The clustering coefficient measures the extent to which a student's connections form tightly-knit clusters or communities. It quantifies the likelihood that two students connected to a third student are also connected to each other. The clustering coefficient of a node i is given by:

$$C_C(i) = \frac{2e_i}{k_i(k_i-1)} \tag{2}$$

where e_i is the number of connections between the neighbors of node i, and k_i is the degree of node i.

4.3 Betweenness Centrality

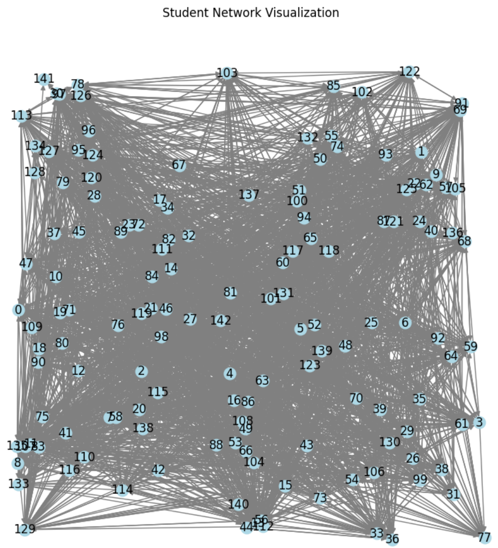
Betweenness centrality measures the importance of a student as a bridge or intermediary between different groups within the network. It quantifies the number of shortest paths that pass through a particular node. The betweenness centrality of a node i is defined as:

$$C_B(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{3}$$

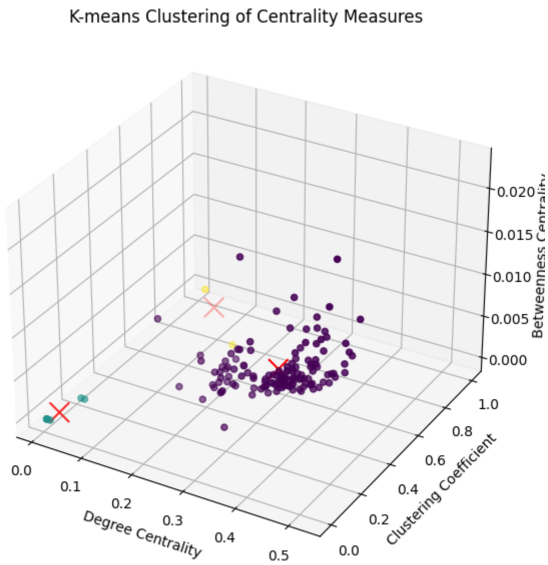
where σ_{st} is the total number of shortest paths from node s to node t, and $\sigma_{st}(i)$ is the number of those paths that pass through node i.

5 Network Visualization and Clustering Results

Figure 1a shows the student network visualized using a random layout. Figure 1b illustrates the results of the K-means clustering algorithm applied to the centrality measures.



(a) Student Network Visualization with Random Layout



(b) K-means Clustering of Centrality Measures

6 Algorithm for At-Risk Student Identification

To identify students who may be at risk of academic difficulties based on their network connections, we propose the following algorithm:

Algorithm 1 At-Risk Student Identification

Require: Student network G, degree centrality C_D, clustering coefficient C_C, betweenness centrality C_B
Ensure: List of at-risk students

Compute centrality thresholds t_D, t_C, t_B using clustering techniques (e.g., K-means)
$at\_risk\_students \leftarrow \emptyset$
for each student i in G **do**
  if $C_D(i) < t_D$ **and** $C_C(i) < t_C$ **and** $C_B(i) < t_B$ **then**
    $at\_risk\_students \leftarrow at\_risk\_students \cup \{i\}$
  end if
end for
return $at\_risk\_students$

The algorithm takes the student network G, along with the computed centrality measures (degree centrality C_D, clustering coefficient C_C, and betweenness centrality C_B) as input. It first computes the centrality thresholds t_D, t_C, and t_B using a clustering technique, such as K-means. These thresholds are used to identify students with low centrality values.

Next, the algorithm iterates through each student i in the network and checks if their centrality measures are below the respective thresholds. If a student's degree centrality, clustering coefficient, and betweenness centrality are all below the corresponding thresholds, they are considered at risk and added to the *at_risk_students* set.

Finally, the algorithm returns the *at_risk_students* set containing the nodes (students) identified as potentially at risk of academic difficulties based on their network connections.

7 Conclusion

Network analysis offers a powerful paradigm for understanding the complex social and academic dynamics within educational institutions. By leveraging the intricate connections and relationships among students, we can gain valuable insights into academic performance and potential challenges. The proposed methodology has the potential to revolutionize the way educational institutions approach student support, enabling proactive measures and targeted interventions to foster academic success and social integration. Through collaborative efforts between researchers, educators, and student support services, this approach can contribute to creating inclusive and supportive learning environments that prioritize the well-being and success of every student.